PROTEIN CHEMISTRY
'DESIGNING' BRAND NEW PROTEINS

STRUCTURAL BIOLOGY
AT THE UK'S LARGEST PARTICLE ACCELERATOR

50 YEARS OF BIOCHEMICAL BREAKTHROUGHS

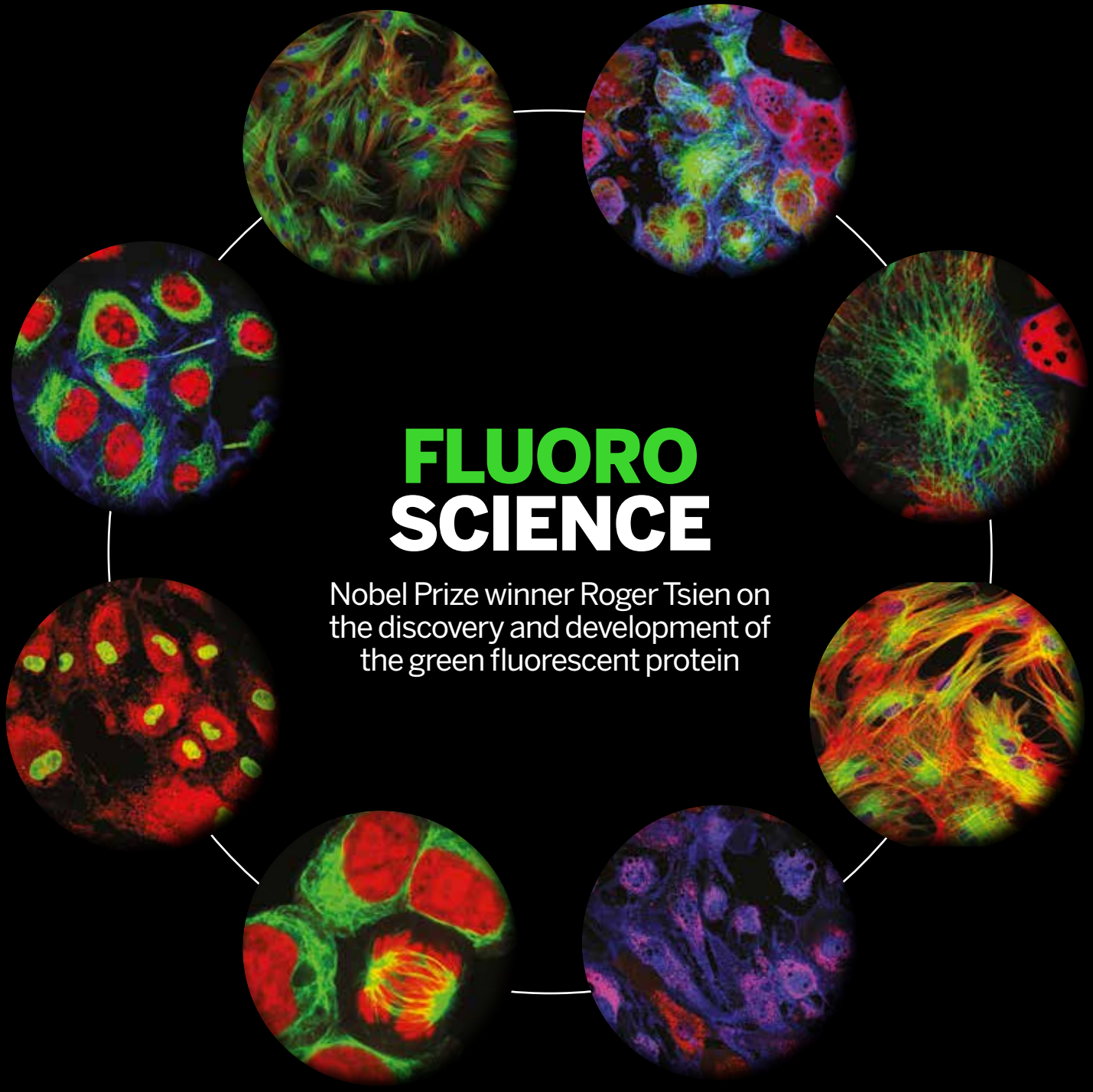
The Biologist:

BIOCHEMICAL SOCIETY

A BIOCHEMISTRY SPECIAL

THE MAGAZINE OF THE ROYAL SOCIETY OF BIOLOGY / www.rsb.org.uk

Vol 62 No 5 supplement • Oct/Nov 2015



FLUORO SCIENCE

Nobel Prize winner Roger Tsien on the discovery and development of the green fluorescent protein

Royal Society of Biology

Contents

Volume 62 No 5 Oct/Nov 2015

Royal Society of Biology

ROYAL SOCIETY OF BIOLOGY
Charles Darwin House,
12 Roger Street, London WC1N 2JU
Tel: 020 7685 2550
Fax: 020 3514 3204
info@rsb.org.uk
www.rsb.org.uk

Managing editor
Tom Ireland MRSB
@Tom_J_Ireland
tom.ireland@rsb.org.uk

BIOCHEMICAL SOCIETY

BIOCHEMICAL SOCIETY
Charles Darwin House,
12 Roger Street, London WC1N 2JU
Tel: 020 7685 2400
Fax: 020 7685 2467
info@biochemistry.org
www.biochemistry.org

Science editor, The Biochemist
Frederica Theodoulou
freddie.theodoulou@
rothamsted.ac.uk

Community and press editor, Biochemical Society
Helen Albert
helen.albert@biochemistry.org

The Biologist and its supplements are produced on behalf of the Royal Society of Biology by

Think Publishing Ltd
Capital House
25 Chapel Street
London NW1 5DH
www.thinkpublishing.co.uk
020 3771 7200

Art director Matthew Ball
Designer Dominic Scott
Production editor Sian Campbell
Sub editor Kirsty Fortune
Publisher John Innes
john.innes@thinkpublishing.co.uk

The Biochemist

This is an exciting time to be a biochemist as new tools and technologies are offering unprecedented insights into the molecular workings of life. It has been a pleasure to help to develop this special issue for *The Biologist* and share a small sample of what biochemistry has to offer with a wider audience. If you like what you see in this taster issue, please do take a look at *The Biochemist*, to explore the wonderful world of molecular biosciences.

Professor Frederica Theodoulou, science editor, *The Biochemist*

About this issue

Produced in partnership with the Biochemical Society, this mini special issue aims to showcase the fascinating and important work that biochemists do. Biochemistry has created the tools, techniques and knowledge on which modern bioscience depends. It is at the very core of how all life on Earth works and its output underpins many other areas of the life sciences including medicine, pharmaceuticals, agriscience and biotechnology. Our content here is focused towards the exciting chemistry of proteins, just one of many different types of macromolecule studied by biochemists. We can never do justice to such a remarkable field in just 16 pages, but we hope this one-off biochemistry special whets your appetite to find out more.

Tom Ireland, managing editor, Royal Society of Biology

IN THIS SUPPLEMENT

TWITTER twitter.com/BiochemSoc

FACEBOOK www.facebook.com/biochemicalsociety

WEBSITE www.biochemistry.org

Roger Tsien on the development of the green fluorescent protein
PAGE 6

Breakthroughs in biochemistry at the Diamond Light Source
PAGE 12

Welcome

Biochemistry is sometimes compared to cookery. Chefs and biochemists both mix ingredients and wait with excited expectation for the result – either a delicious new dish or an experimental outcome. They both follow recipes, although biochemists refer to theirs as 'experimental protocols'. Continuing this metaphor, I am reminded of the famous recipe for jugged hare that is said to start with "first catch your hare".

Biochemists do not normally race around the countryside chasing furry animals, but until recently there was a parallel preliminary step in our protocols. We had to process litres of culture or extract kilograms of tissue before we could start work with milligrams of the molecule of interest.

We now operate on a micro scale way beyond the tiniest amuse bouche in a nouvelle cuisine restaurant. We can start with tiny amounts of tissue and get information about thousands of molecules, rather than just one as in the past.

With genomics and molecular biology, we can isolate genes affecting processes that were previously inaccessible to the biochemist. From the genes we see the proteins, and from the proteins we find other components of the biochemical circuitry in the cell or organism.

> We are 'bio' chemists and our ultimate goal is to understand how living systems are more than the sum of their parts, for the benefit of humankind

Professor David Baulcombe, president of the Biochemical Society

Metaphors normally break down under close inspection and this one is no exception. Few people would compare modern molecular genetics to mere cookery (and biochemists certainly do not 'cook' their results...). There is, however, one element of the cookery metaphor that still applies: slow food.

Slow food enthusiasts would relish catching the hare, and they may embrace new technology, but they do not wish to lose sight of the whole food chain. Biochemists need to remember this slow food movement as we bury ourselves in the enormous amounts of data pouring out of our mass spectrometers, imaging devices and next generation sequencers.

We are 'bio' chemists and our ultimate goal is to understand how living systems are more than the sum of their parts, for the benefit of humankind. We should not lose sight of the biological hare that is our raison d'être. Are we succeeding? I believe we are. Readers of *The Biologist* can judge for themselves by reading this special issue on the field.

50 YEARS OF BIOCHEMISTRY

Seven leading biochemists pick the most important European breakthroughs of the past five decades*

DNA sequencing

In this method of DNA sequencing, chemically altered nucleotides terminate newly synthesised DNA fragments at specific bases – either A, C, G or T. These fragments are then ranked by size, and the DNA sequence can be read by detecting which altered nucleotide is present at each different length fragment.

LOCAL WORLD/REX SHUTTERSTOCK

I have to choose Fred Sanger for the discovery that DNA can be sequenced by a clever technique known as the chain termination method, or Sanger sequencing.

The development of this method came hard on the heels of the Maxam–Gilbert method from Harvard, which was based on chemical cleavage of DNA and was rather hard work. The amazing thing about the Sanger method is that it was so elegant, so robust and so simple to use. Of course, in time it became automated, which led to the human genome project.

Sanger's method remained the universally adopted and undisputed best way to sequence DNA for nearly 25 years, and it's only in the past decade that it has been supplanted by myriad new higher throughput methodologies. It had an impact on so many areas of our thinking, including simplifying and unifying genetics, the discovery of the third kingdom of life, understanding how DNA evolves, the importance of DNA modifications, and the rise and rise of bioinformatics as an essential experimental approach to any problem.

Professor Stephen Busby, University of Birmingham, UK

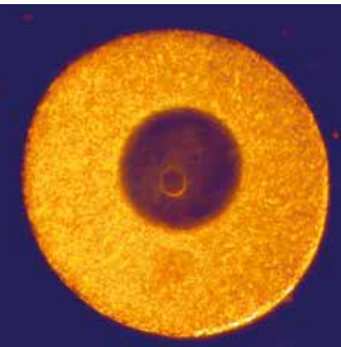
Science luminaries, from left: James Watson, Max Perutz, Cesar Milstein, Fred Sanger, John Kendrew and Aaron Klug

The amazing thing about the Sanger method is that it was so elegant, so robust and so simple to use

*This article is an edited version of '50 years of European Biochemistry', first published in *The Biochemist*, August 2014

The cell cycle

Multiple checkpoints in the eukaryotic cell cycle ensure that cells only divide after sufficient growth and faithful DNA replication – a process essential to preventing cell division going awry. Of the many proteins involved in cell cycle control, cyclin-dependent kinases (CDKs) are among the most important, modifying other chemicals involved in the cell's progression towards division.



MICHEL DELARUE, ISM/SCIENCE PHOTO LIBRARY

I remember as a PhD student back in the 1980s making a brief visit to Jim Maller's laboratory in Denver, en route to a meeting in Colorado. Maller's group had discovered a protein from *Xenopus* oocytes that had a key role in controlling cell division. At the time, this seemed a million miles away from my own project – working on a protein involved in regulating lipid metabolism in rat liver – but the significance of the work was not lost on me.

Shortly afterwards, I heard about work from Tim Hunt's group at the Imperial Cancer Research Fund's Clare Hall laboratories. A small family of proteins had been detected in sea urchin eggs whose levels went up and down synchronously with each cell cycle – proteins that would later be called cyclins.

Sometime later, as a postdoc in Dundee, I heard a talk from Paul Nurse in which he described the identification of a protein from yeast that was regulated by binding to cyclins, and which was required for cell division. Of course, 25 years on, we now know many of the intricate and exquisite details of how the cell cycle is regulated: cyclin-dependent protein kinase (CDK1), encoded by the cdc2 gene in yeast, and together with cyclin B, forms a protein kinase complex known as MPF – the very one that had previously been isolated in *Xenopus* oocytes.

Tim Hunt and Paul Nurse, together with Lee Hartwell, shared the Nobel Prize in Physiology or Medicine in 2001 for their discovery. Hindsight is a wonderful thing, and looking back it is easy to see how all the pieces slotted together, but at the time this was far from the reality. This simply reflects the nature of most scientific discoveries – very rarely are things crystal clear in the heat of the moment, and cut and dried results tend to be the exception rather than the rule. Nonetheless, the discovery of the fundamental mechanisms regulating the cell cycle is a beautiful example of the elegance of nature itself.

Professor David Carling, Imperial College London, UK

The chemiosmotic theory

Peter Mitchell's chemiosmotic theory illustrated how the movement of ions through biological membranes could provide useful energy to catalyse biological processes. Most famously it demonstrated that the movement of hydrogen ions through mitochondrial membranes enabled cells to create the energy carrying molecule adenosine triphosphate (ATP).

If it wasn't for the beauty of this theory that graced my undergraduate lectures and textbooks 30 years ago, my career might have taken a very different step. The theory showed me that biochemistry could be elegant as well as functional.

In the 1950s and 1960s, enzymes did two-dimensional chemistry, and cells and organs did three-dimensional biology. The chemiosmotic theory used physics to bridge these dimensions, uniting chemistry and biology. And it did it for one of the most important biochemical questions: how do we efficiently convert the potential energy from the food we eat and the oxygen we breathe into a useful form that can power movement, development and reproduction?

This theory beautifully explained disparate strands of theory and confusing experiments, converting the mess in this field in the 1960s into the clarity of the 1980s. With our 21st century world view of mitochondria shaped by high resolution graphics and conceptual video reconstructions, it is difficult to understand how truly paradigm shifting the chemiosmotic theory was. The first understanding of membrane protein structure was still more than 20 years away – they were known to associate with lipid bilayer membranes, but were not considered to be integral to their structure.

After the theory, a new concept of membrane enzymes as multidimensional entities emerged, equally as able to catalyse the transport of ions as form chemical bonds, and indeed it was found they did both simultaneously. The theory became so famous that it was used as an example of paradigm shifting in sociology books.

Even after the award of the 1978 Nobel Prize to Mitchell, there were still dissenters. However, by the dawn of the new millennium, the world, and its textbooks, had accepted the theory.

Professor Chris E Cooper, University of Essex, UK


Rosalyn Yalow

Antibody technologies

Where would we be without them?

When I show visitors around my laboratory they are always amazed that scientists can make any conclusions from an experiment performed in a few microlitres of clear liquid or from looking at little black bands on a computer screen. Of course, one of the main reasons we can do such experiments is the power of the humble antibody.

Proteins do most of the work in any cell, and their activities and their contributions are ever changing as the levels of individual proteins rise and fall, their sub-cellular localisation changes and post-translational modifications alter the way they work.

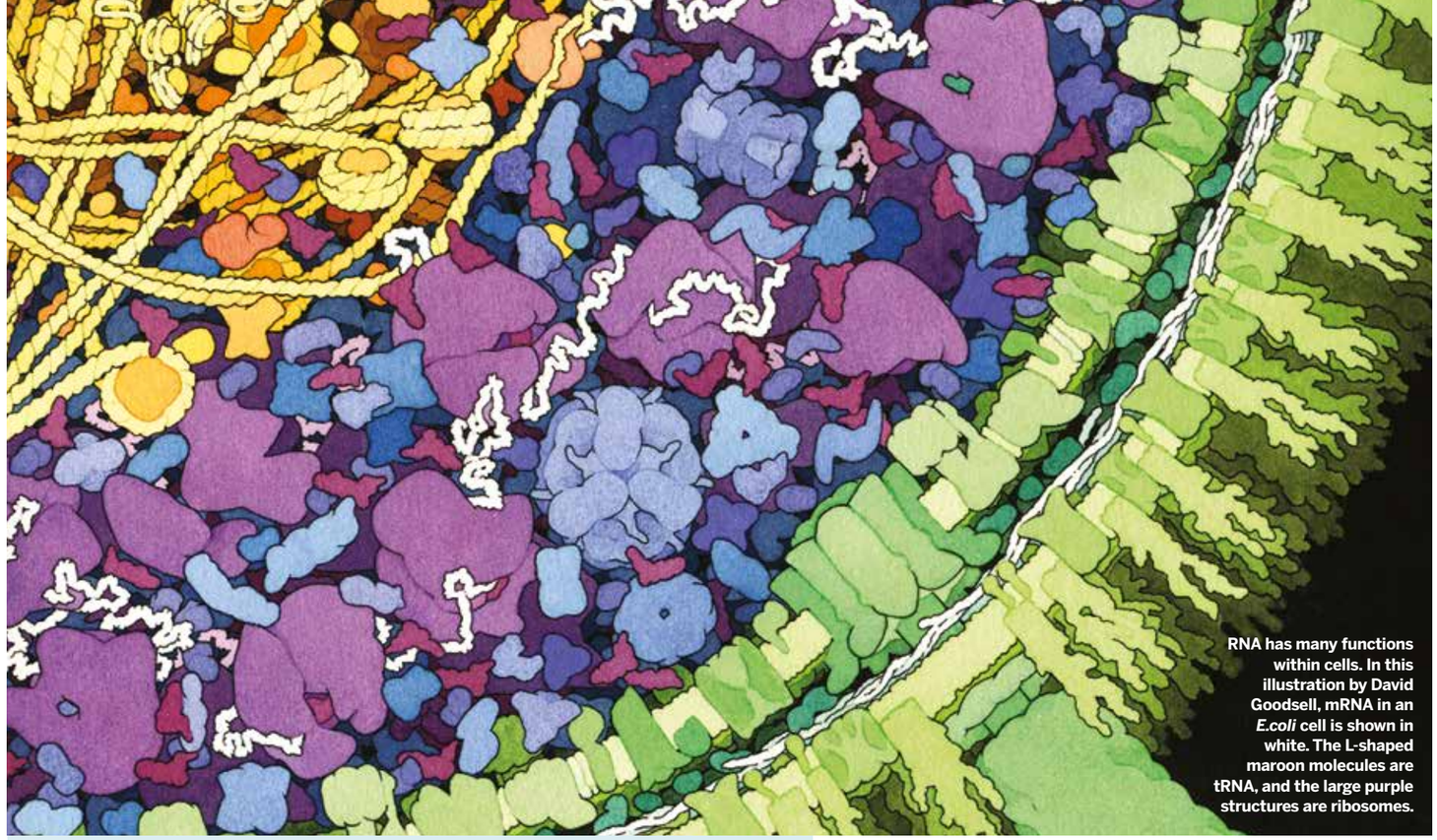
In the 1960s the study of proteins was in large part reliant on tedious purification protocols. Now, there are tens of thousands of well characterised antibodies we can use to identify and study proteins easily. Although we may debate the relative merits of polyclonal and monoclonal antibodies, both reagents are essential for modern biochemical research.

The first large scale use of antibodies in biochemistry was facilitated by the development of the radioimmunoassay by Rosalyn Yalow and Solomon Berson, which allowed biochemists to measure concentrations of antigens with extreme precision. Later perfected by Charles Nicholas Hales and Philip Randle, this revolutionised the quantitation of biological molecules, but the use of radioactive isotopes limited the broad use of these assays.

Techniques that use fluorescently labelled antibodies, together with methods that could link antibodies to colloidal gold for electron microscopy, have become the mainstays of modern cell biology.

Advances in antibody based technologies in the last 20 years provide us with a comprehensive set of tools to assist in the analysis of virtually any biological function in the cell. Without them, modern biological research would not exist as we know it.

Professor Peter Shepherd, University of Auckland, New Zealand



RNA has many functions within cells. In this illustration by David Goodsell, mRNA in an *E.coli* cell is shown in white. The L-shaped maroon molecules are tRNA, and the large purple structures are ribosomes.

NATIONAL LIBRARY OF MEDICINE/SCIENCE PHOTO LIBRARY; ILLUSTRATION BY DAVID S GOODSELL, THE SCRIPPS RESEARCH INSTITUTE

The discovery of small interfering ribonucleic acids

As well as its role in protein formation, ribonucleic acid (RNA) can play a key role in gene regulation and protecting against pathogens.

Until around 15 years ago, it was thought there were two main roles of RNA in cells. First, it can be a structural molecule – for example, acting as a scaffold for building the ribosomes that carry out protein synthesis. The second and perhaps the best understood role of RNA in the cell is its functional role as messenger RNA (mRNA): DNA is transcribed to mRNA, which is translated to protein.

The view of RNA as mainly an intermediary molecule in gene expression began to be dismantled in the 1990s when small non-coding RNAs, an entirely new category of RNA, were discovered.

For many years, double-stranded RNA (dsRNA) had been recognised as an important part of the innate immune response to virus infection: plant and animal cells detect virus infection by recognising dsRNA species produced during virus replication. Cells have evolved mechanisms to detect and degrade foreign dsRNAs as a means to inhibit successful viral replication.

David Baulcombe, together with Andrew Hamilton, working at the Sainsbury Laboratory in Norwich, discovered the presence of extremely small RNA molecules that mediated the phenomenon of this RNA silencing in plant cells. The average messenger RNA is 2,000 bases in length but the average size of these so-called short interfering RNAs (siRNA) is only 21–24 bases long.

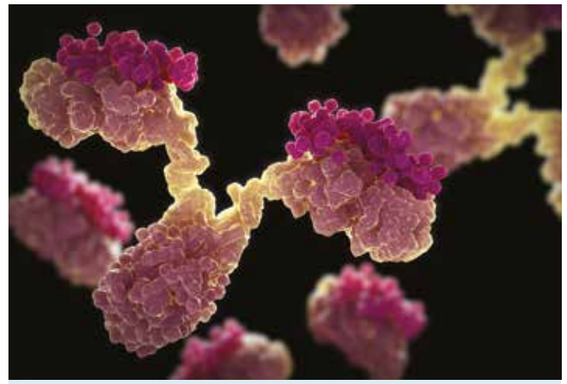
The second surprise was that these ubiquitous ultra-short plant RNAs are not only important in defence against viruses and mobile genetic elements, but they can also regulate developmental gene expression. In other words, these short RNAs have essential functions in cells.

Crucially, the discovery of RNA interference (RNAi) has provided essential experimental tools for the dissection of biochemical pathways in cells. More recently, RNAi has been used in more than 30 clinical trials targeting genetic diseases, cancers and viral infections.

The decision by a group of plant biologists to look at these tiny RNA species has seeded an entirely new field of biochemistry. And what would today's laboratory bioscientist do without the means to manipulate gene expression using small interfering RNAs?

Professor Sheila Graham, University of Glasgow, UK

Below: Human antibodies, the Y-shaped proteins used by the immune system to defend against foreign objects like bacteria and viruses



Monoclonal antibodies and their exploitation

Monoclonal antibodies are an essential part of any molecular biologist's toolkit. Antibodies that bind specifically to almost any substance can now be produced and used to detect, target or purify that substance.

The invention of monoclonal antibodies by César Milstein and Georges Köhler, followed by their adaptation for medical exploitation by Greg Winter and colleagues, has had an enormous impact on everything from diagnostic pathology, protein purification, gene isolation and countless aspects of cell biology, to direct therapeutic treatment for major diseases such as cancer. It has spawned a multibillion-pound industry and resulted in the extraordinary statistic that approximately a third of all recently approved therapeutic drugs have been modified monoclonal antibodies.

In the mid-1970s, a crucial question was: how does the immune system produce such an extraordinary range of diverse antibodies? Milstein's group was looking for a stable reproducible source of antibody production that would help them study such questions.

Myeloma cells were immortal and continued to produce antibody in culture, but the antibodies were weak and unsuitable for research. Milstein and Köhler overcame these problems by fusing antibody-producing spleen cells to immortal myeloma cells, producing hybrids, each of which produced a single antibody indefinitely.

This resulted in the classic 1975 *Nature* paper, which ended with the profound insight that "such cultures could be valuable for medical and industrial use", a classic understatement, as Milstein and Köhler were awarded the Nobel Prize for Physiology or Medicine in 1984.

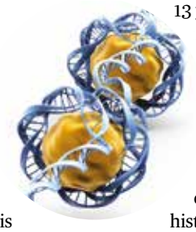
Since that time, monoclonal antibodies have been exploited in extraordinary ways that Milstein and Köhler could not have expected. These range from gene isolation to purification of proteins and other cellular components, immunofluorescence, and an extraordinary range of immunodiagnostic tools and approaches.

Professor Ron Laskey, University of Cambridge, UK

Solving the nucleosome structure

How is the entire genome packed into the nucleus?

The nucleus has a big problem. In a typical human cell, the nucleus measures just 5–10μm in diameter, yet contains 6.4×10^9 base pairs of DNA. That's over two metres' worth of DNA crammed into this small space. The solution to this is packaging (pictured). The DNA is wrapped up by a set of histone proteins into nucleosomes, which are themselves arranged into higher ordered structures.



ALFRED PASIEKA/SCIENCE PHOTO LIBRARY/SCIENCE PICTURE CO/SCIENCE PHOTO LIBRARY

Work to understand the structure of the nucleosome at the atomic level began in earnest in the 1980s in Aaron Klug's laboratory at the MRC Laboratory of Molecular Biology in Cambridge. It was already known that the nucleosome core particle consists of two copies each of the histones H2A, H2B, H3 and H4. This forms an eight histone structure around which roughly 150 base pairs of DNA are wrapped.

In an x-ray crystallography tour de force, Timothy Richmond and his colleagues solved the structure of the nucleosome–DNA complex at near atomic resolution some 13 years later. At this resolution, the path of the DNA helix as it encircled the histone octamer could be traced. The structure also revealed that the ends of some of the histones protrude out, forming tails that can interact with nearby nucleosomes to create higher-order folding arrangements. These modifications have enormous biological significance in the regulation of gene expression.

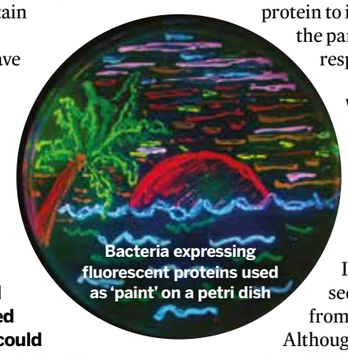
It might be easy to think that the two metres of DNA in each nucleus is like a ball of string. It is, however, far from it. The more we understand, the more we appreciate why things have to be the way they are.

Professor Richard J Reece, University of Manchester, UK

The light fantastic

The remarkable American biochemist Roger Y Tsien tells us how he made his most famous discovery

In 2008 Roger Y Tsien shared the Nobel Prize in Chemistry for the discovery and development of the green fluorescent protein (GFP). This glowing molecular tool has revolutionised many areas of biochemistry research, allowing researchers to visualise the expression of certain genes or certain molecules within cells. Molecular biologists have since found countless uses for GFP and similar molecules, and fluorescent proteins are now an essential part of biochemists' molecular toolkit.


Bacteria expressing fluorescent proteins used as 'paint' on a petri dish

Before you discovered GFP, your work involved looking for dyes that could help image neuronal activity. What inspired you to work in this field?
The visual system is the only sensory system with the ability to display lots of events in spatiotemporal detail, so one has to use one's own visual system to investigate another creature's nervous system. From very early on in graduate school, I was attracted to developing techniques for visualising neuronal activity as the best way to resolve many neurons firing simultaneously.

What led you to look at fluorescent proteins and their related genes?
My colleagues and I had painstakingly built dyes such as Fura-2 and Indo-1 – with molecular weights near 840 – for recognising and visualising small calcium ions, whose molecular weight was only 40. So it seemed that for the more general problem of recognising biochemical messengers such as cyclic AMP (molecular weight 329), let alone full sized proteins, we would have to adopt the techniques of molecular biology rather than synthetic chemistry.

I started in around 1988 by discussing a collaboration with Alexander Glazer on phycobiliproteins, a family of fluorescent proteins from blue-green algae. However, these needed a separate partner protein to insert the chromophore, the part of the molecule responsible for its colour.

Why was the fluorescent protein of the jellyfish *Aequorea victoria* so useful?
In 1992, Douglas Prasher at the Woods Hole Oceanographic Institution cloned and sequenced the gene for GFP from *Aequorea victoria*.

Although he was unable to work on GFP any further himself, he was willing to give samples of its DNA to requestors, of which there were two: Martin Chalfie and me. Marty's lab discovered that GFP didn't need help from any other protein in the jellyfish, so GFP had both availability and autonomy. It has taken us almost 30 more years to engineer an easily expressible phycobiliprotein.

I was attracted to developing techniques for visualising neuronal activity as the best way to resolve many neurons firing simultaneously

SIPA PRESS/REX SHUTTERSTOCK

© WENN LTD/ALAMY

Left to right: Nobel Prize winners Paul Krugman, Martin Chalfie and Roger Tsien with the then US president George W Bush

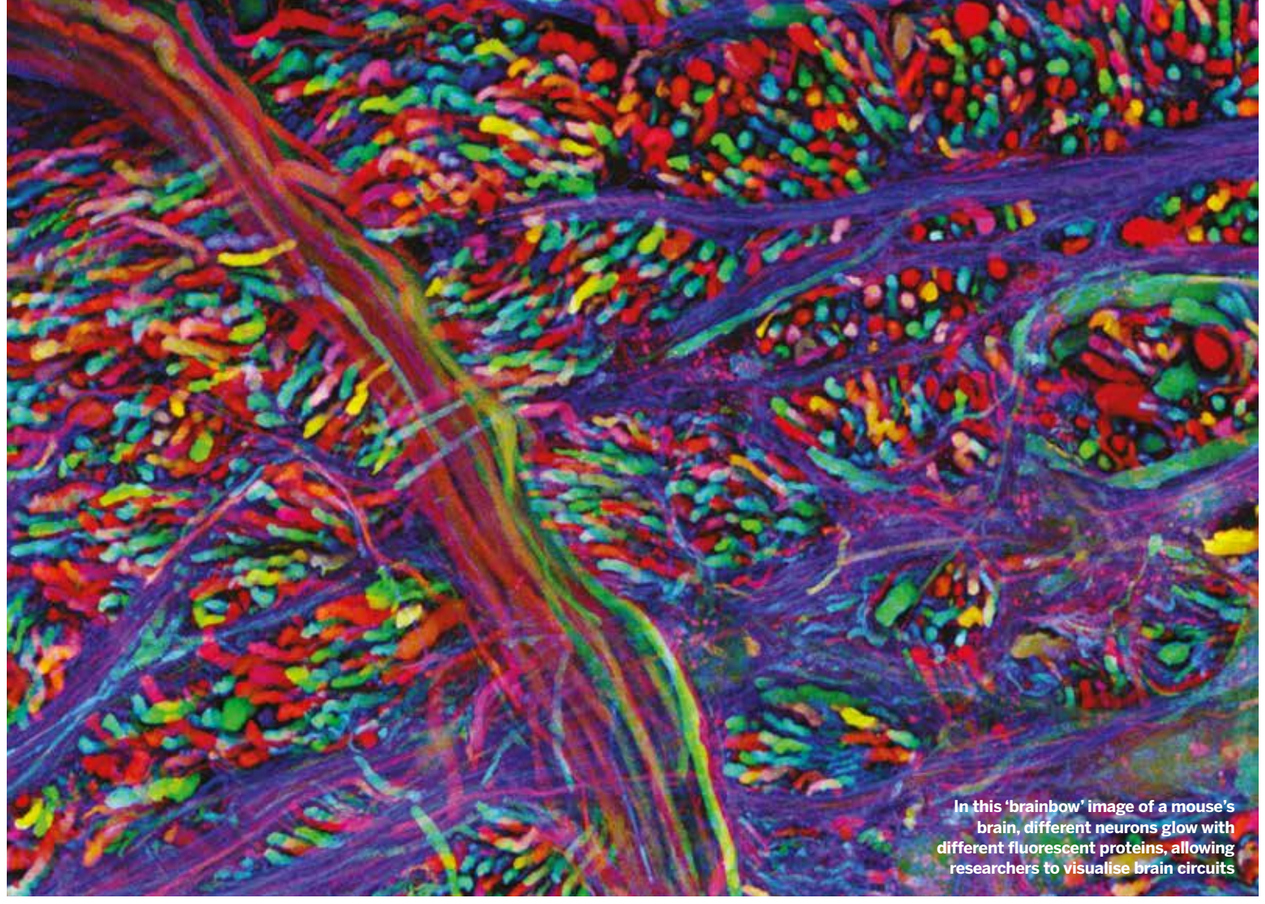
Did you ever imagine that GFP and its derivatives would be used by so many researchers in so many different ways?
I knew that an autonomously fluorescent protein module would be of immense value, but I didn't anticipate it would have quite so many uses.

Do you have a favourite way in which GFP has been used?
It was satisfying when we got a phenomenon called fluorescence resonance energy transfer (FRET) working between mutants of GFP. FRET senses the proximity of two fluorophores of different colours and had been a major goal when we set out. But that's now long in the past.

How else do you think fluorescence might be used in the future?
I can't foresee a limit to future applications of fluorescence. After all, fluorescence is an unusual and very useful property of a small proportion of molecules. Under the right circumstances it can be observed in anything from single molecules to oceans, over nanoseconds to many days, using the naked eye to the most sophisticated instruments.

Can you tell us a little about fluorescence assisted cancer surgery?
In cancer surgery, fluorescence guidance would be helpful because tumour tissue doesn't look any different from normal tissue under ordinary white light illumination. We want to use biochemical differences between the tumour and normal tissue to make the tumour fluorescent, so that the surgeon can decide where to cut with realtime guidance.

Unfortunately, one cannot use GFP or its homologues, because they can be linked to malignancy only by sophisticated gene therapy that is not practical yet or ethical in human patients. Instead, we are exploiting extracellular enzymes that are turned on in practically all solid malignant tumours. We have engineered fluorescent substrates that are triggered by these enzymes to enter cells and become trapped, and also to change colour by modulating FRET (the same phenomenon mentioned above). A small biotech company partly founded by me has just started a clinical trial with such molecules, together with the instrumentation for surgeons to see the fluorescence as they operate.



In this 'brainbow' image of a mouse's brain, different neurons glow with different fluorescent proteins, allowing researchers to visualise brain circuits

© LISA WERNER/ALAMY STOCK PHOTO

What is GFP?

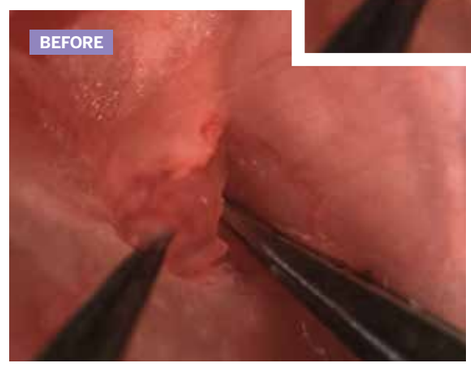
GFP stands for green fluorescent protein. It is a protein that glows green in the presence of UV or blue light, originally found in the bioluminescent and fluorescent jellyfish *Aequorea victoria*.

In 1992, the gene for GFP was sequenced by American biologist Douglas Prasher. The first to express the gene in another organism was Martin Chalfie, an American biochemist who shared the Nobel Prize with Tsien. He inserted the gene for GFP into the bacteria *E. coli* and nematode worm *C. elegans*. The resulting organisms then glowed green in the presence of UV or blue light.

The protein itself is a barrel shaped molecule with a colour producing 'chromophore' in the centre – it is thought just three amino acids in the protein chain create the fluorescent 'chromophore'. It is stable, non-toxic to most organisms when expressed in cells, and requires only UV/blue light and oxygen to emit its eerie glow, making it perfect for *in vivo* applications.

Green fluorescent protein has since been used in thousands of different ways. Replacing a gene with the gene for GFP can result in GFP being expressed in the organism only in the places where the original gene would have been expressed, creating a bright visual pattern of expression. By selectively labelling specific proteins, we can create images to see exactly where those proteins are present.

We want to use biochemical differences between the tumour and normal tissue to make the tumour fluorescent



LIGHTING THE WAY
A matched pair of photographs showing a tumour about to be excised, viewed without and with the aid of tumour imaging peptides

DR JEAN LIVET/WWW.OLYMPUSBIOSCAPES.COM

What else is your lab working on at the moment?
We are trying to gather evidence for a new hypothesis for how and where the brain might store permanent memories at the molecular level[1].

Previous hypotheses have assigned the site of memory storage to be various proteins within synapses, the places at which neurons communicate with each other. The difficulty with these hypotheses is that proteins inside synapses undergo continuous rapid turnover and replacement, so that memories would require recopying very many times over an animal or person's lifetime.

Instead, we are looking at the glycoproteins (proteins plus carbohydrates) known to form a coating just outside synapses. We are accumulating evidence that this coating, once formed, is basically stable but can be locally remodelled to strengthen individual synapses – and thus serve as molecular substrates for memory.

You hold around 100 or so patents for various other biotechnology tools. Which are you most proud of?
In 1994, we started a biotech company called Aurora Biosciences to use new fluorescence assays to speed up drug screening in the pharmaceutical industry. One of the projects Aurora took on was to find drugs to help cystic fibrosis. Most experts thought Aurora's chances were negligible, as the market for such rare disease remedies was thought to be too small, and gene therapy was considered a much more promising approach. However, the Cystic Fibrosis Foundation backed Aurora's efforts, and fluorescence screening found the drug that was recently lauded by President Obama as an example of 'precision medicine'. Such a long time is required before one knows whether one has success or not.

Were you interested in science as a child?
I was always obsessed by pretty colours and by technologies that seem useful. One of my earliest memories is of a beach that had a zone of coarse pebbles surrounded by two zones of sand. I tried to lay down a bridge of sand across the pebbles to make crossing more comfortable for my tiny bare feet. Of course, the bridge would have been washed away by the next big wave or high tide. Perhaps that's a metaphor for much of my career.

You have a long tradition of engineering in the family. Do you consider yourself a chemist, biologist, bioengineer or what?
I'm a muddled mix. When I was applying for my first faculty position, several biology departments rejected me on the grounds that I was a chemist, and at least one chemistry department turned me down as too much of a biologist. Almost all my work has been involved with tool building, but I have never had a formal engineering course or appointment. Fortunately, most forward looking departments have now adopted a more flexible and interdisciplinary viewpoint. Personally, I don't care much for labels.

References
1) Tsien, R. Y. Very long-term memories may be stored in the pattern of holes in the perineuronal net. *Proc. Natl. Acad. Sci. USA* **110**(30), 12456–12461 (2013).

Roger Y Tsien is professor of pharmacology and professor of chemistry and biochemistry at the University of California, San Diego. After graduating from Harvard, Tsien also held posts at Cambridge and Berkeley. He is also a noted biochemical inventor who holds more than 100 patents. He shared the Nobel Prize in Chemistry in 2008 for his development of GFP.

Little green chemists

Understanding the chemistry of plants will enable us to create hardier and more useful crops, writes Joseph Jez

Plants are the original master chemists. They create a dizzying array of small molecules for all sorts of purposes: harvesting light; making nutrients and cellular building blocks; fighting off pathogens, insects, herbivores and sometimes each other; attracting pollinators; communicating with soil microbes or other plants; stocking seeds for the next generation; and adapting to a changing environment. Growing where their seed lands, plants are the consummate DIYers and use their metabolic arsenals to thrive and succeed.

Unravelling plants' molecular bounty and understanding their metabolism offers new opportunities and lessons for chemists, biochemists, geneticists and ecologists.

My research group focuses on understanding the molecular machinery that generates the molecules used by plants to grow and meet environmental challenges. Understanding these systems could help us engineer plant metabolism to maintain food production under environmental stresses such as drought, or for the production of fuels, chemical feedstock molecules and other useful materials.

A good example of this is the protein known as adenosine 5'-phosphosulfate kinase, which allows plants to use sulphur from the soil in the synthesis of various molecules.

Many environmental stresses, including exposure to heavy metals, pollution, drought and temperature changes, generate reactive oxygen molecules. These molecules can perturb cellular physiology and metabolism to affect plant growth and development. They can also modify the amino acids on a protein. This is what happens to adenosine 5'-phosphosulfate kinase.

Following modification, the protein's structure is changed to slow the chemistry it performs, altering how much product it makes. The structural change in effect becomes an on/off switch for the enzyme that is toggled by the reactive oxygen molecules produced by environmental stress.

These seemingly subtle atomic level differences have larger metabolic consequences for the plant. Slowing down the enzyme lets its substrate be used to make other molecules that help protect the plant against the reactive oxygen molecules.

NO MORE 'GRIND TO FIND'

Ultimately, understanding how such proteins change in response to stresses can provide insights for modifying plants to meet challenges such as maintaining crop production during climate change, or under growing conditions that are less than ideal.

Once, the mantra of plant biochemists was 'grind and find', but the 21st century has brought us new tools. The powerful traditional combination of protein purification, enzyme assays, synthesis of substrates and products, and other molecular biology techniques are fundamental to both understanding and manipulating plant metabolism. And the experimental palette is widening.

The availability of bioinformatics tools and genomic data, the relative ease and low costs of whole genome and RNA sequencing, the ever improving sensitivity of mass spectrometry, synchrotron radiation sources for protein structure determination (see page 14), new imaging reagents and technology, and a growing variety of genetic manipulation technologies are opening new avenues for deciphering how plants work.

TIMELESS FASCINATION

New technologies and changing perspectives blur the lines between protein chemistry, molecular and cellular biology, computer science, genetics and systems biology, but that blending is important because new science comes from the cracks between the disciplines.

The possibility of engineering plant metabolism for bio-based production of fuels, chemical feedstock molecules and materials, while maintaining food production under environmental stresses such as drought, motivates both basic and applied research worldwide. However, it also requires scientists who can move freely from atoms to ecosystems.

Our fascination with plants and explorations of their chemical diversity began in the pre-dawn of our history as a species – the hunter gatherer looking for the right nuts, fruits and leaves to eat, mixing and mashing the particular berries for paint, tasting the bark of a tree because of headache, infusing water with special leaves, or finding the best wood for a spear, a tool or a fire. That tradition continued through the advent of agriculture, the creation of paper, transportation and houses, the discovery of many pharmaceuticals, and the harnessing of fossil fuels. Today, the same technologies that advance our understanding of diseases and biomedical applications are also revealing the inner workings of plants and change the way we ask questions in pursuit of a sustainable future.

Joseph Jez is professor of biology at Washington University in St Louis

JON WILSON/SCIENCE PHOTO LIBRARY

Unravelling plants' molecular bounty and understanding their metabolism offers new opportunities and lessons

An x-ray crystallography machine

Protein structure

The key to understanding metabolism

Knowing where nearly every atom is in a protein is our hypothesis generator: how does chemistry happen? Why does the protein recognise a particular metabolite, another protein or a DNA sequence? Can we change the protein structure and does that alter its function? Can we use that change for a practical purpose?

At the heart of those efforts is our use of x-ray crystallography to unveil the three-dimensional structure of proteins. This experimental tool begins with growing crystals of our protein of interest. Next, we place a crystal into a focused x-ray beam. When the x-ray beam hits those atoms in the crystal, the x-rays then bounce off – that is, they are diffracted.

The diffracted x-rays produce a pattern of spots on an x-ray detector. That pattern contains information about the location of every atom in the protein crystal in three dimensions. Using computers to help solve the maths that relates the spots back to the crystal, we generate a map of where the atoms are and reveal what the protein structure looks like.

Shedding light on molecules

Structural biology owes much to the intense beam of the UK's largest particle accelerator

Technology is the cornerstone of scientific progress. Throughout history, developments in equipment, techniques and expertise have all been crucial to supporting discovery and advancing our understanding of the world. Resembling a giant silver ring the size of Wembley Stadium, one of the most iconic examples of technology supporting research is the UK's synchrotron science facility, Diamond Light Source.

Diamond functions like a collection of giant microscopes using x-rays, infrared and UV light.

A third generation synchrotron, or particle accelerator, the machine produces a beam of light at an energy of 3 gigaelectronvolts, which is 10 billion times brighter than the sun. Scientists use it to investigate the atomic and molecular nature of matter.

From viruses and drugs to engineering components and nanotechnology, the UK synchrotron is used for research in virtually all areas and its high-tech capabilities are pushing the boundaries of what is possible for scientists everywhere. However, about 40% of the work that takes place at Diamond is biology related, making it the most popular area of research, and the synchrotron supports a vast range of research in this field, including biomedical studies, bioengineering, environmental science and much more.

Diamond currently has five beamlines dedicated to protein crystallography and a further 19 providing other techniques useful to bioscience, including spectroscopy, small-angle scattering, tomography and powder diffraction, to name a few.

Here are just a few examples of the biochemical breakthroughs being made at the end of the synchrotron beamline.

WATCHING VIRUSES UP CLOSE

Viruses are one of the oldest and most pervasive elements of the natural world, and have been evolving constantly over millions of years. Scientists from Diamond and the University of Oxford are using the synchrotron's high containment virus facilities to chart developments in viruses and their structure over time, building up our knowledge of individual virus families at the atomic level.

This research is helping us to understand more about viruses and the way they evolve – work that will enable us to predict viruses that may emerge in the future. This work could prove particularly important in the event of an epidemic, when knowing a virus's next step could enable us to respond quickly.

Understanding the atomic structure of viruses is also helping in the development of next-generation vaccines. Some researchers working with Diamond have successfully replicated the atomic structure of foot and mouth disease, creating a 'lookalike' empty shell, devoid of the viral component (RNA). The shell vaccine then induces an immune response without any threat of infection. Currently in clinical trials, this method has the potential to yield safer, cheaper and more effective vaccines, and scientists are also looking to use it to develop a new vaccine for polio.

THE IMMUNE RESPONSE IN REAL TIME

A team from Cardiff University is currently using the machine to study T cells, a variety of white blood cell responsible for finding and destroying unwelcome cells. The scientists have been studying receptors on the outside of T cells that help identify cells containing foreign proteins.

Scientists have been able to observe this process in astonishing detail and are now looking to develop ways of enhancing patients' own T cells so that they can recognise and bind to cancerous proteins, destroying tumours.

In separate research, a group from Queen's University Belfast is exploring the potential of gold nanoparticles to enhance radiotherapy. The scientists are using Diamond to study chemical reactions between nanoparticles and surrounding tissue when exposed to radiation. Their research shows that by injecting nanoparticles to the site of a tumour, it may be possible to increase the therapeutic impact of radiotherapy, enabling doctors to use less radiation.

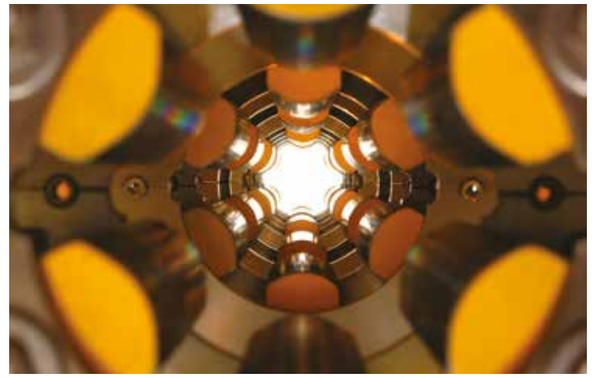
FILLING PROTEINS' POCKETS

Approximately 30% of drugs target G-protein-coupled receptors (GPCRs), making them the largest and most important family of drug targets in the human body. Using the synchrotron, researchers from Heptares Therapeutics have identified the structure of one of the receptors in the brain responsible for the stress response.

By visualising this stress protein receptor at the atomic level, they have identified a 'pocket' in the structure. These findings will enable scientists to design a drug to fit precisely into the pocket, inhibiting the response of the receptor and offering new, more targeted treatment options for anxiety and depression.

MOLECULES AND MALNUTRITION

Away from biomedical research, scientists at Rothamsted Research are using Diamond to look at wheat and its nutritional content. Wheat is one of the most popular foods in the world: combined with rice and maize, it comprises 60% of all food consumed on Earth. The grains are packed with essential nutrients, but much of these are locked away in forms that cannot be processed by enzymes in the human gut. Scientists are using Diamond's spectroscopy capabilities – including elemental mapping, fluorescence and x-ray tomography – to explore methods of changing the way the grains store their nutrients so that they become easier for our guts to access.



Magnets (left) drive the electron beam around the ring of the Diamond Light Source synchrotron particle accelerator. The Oxfordshire facility (below) continues to shed light on how biological molecules function, as well as the structure of viruses.

Diamond currently has five beamlines dedicated to protein crystallography and a further 19 providing other techniques useful to bioscience

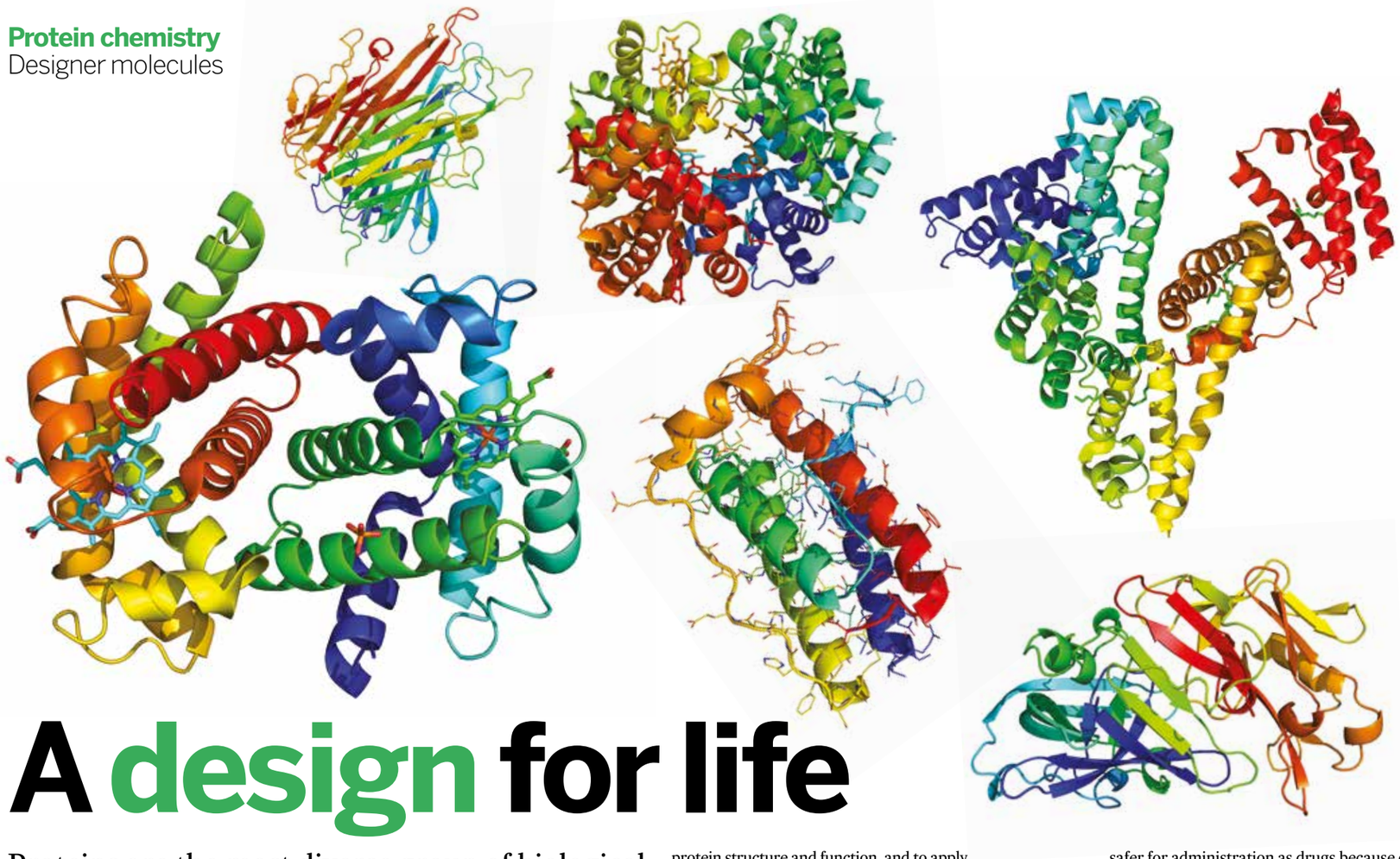
RING OF POWER

Work that once took decades to complete can now be done at the synchrotron within an hour. However, it would be wrong to assume that the field has become routine. Indeed, advanced technology is allowing bioscientists to explore the structures and compositions of bigger complexes and more difficult proteins.

Venki Ramakrishnan, president-elect of the Royal Society, was awarded the Nobel Prize in Chemistry in 2009 for solving the structure of the ribosome, the large molecular machine that strings amino acids together to create proteins. Not too long ago, this profoundly complex biological component would have been impossible to study in any detail. However, advances in technology have opened up more sophisticated techniques, enabled scientists to collect data more quickly, and made it possible to explore more deeply into cells than ever before.

Commenting in a Royal Society paper celebrating Diamond's scientific achievements, Ramakrishnan said progress in structural biology in the last two decades had been "truly remarkable. In large part, this has been due to x-ray crystallography using synchrotron radiation".

Bioscience is a rapidly evolving field and this is also true for the technology that supports it. Facilities such as Diamond are constantly upgrading to allow scientists to do more. Things certainly move incredibly quickly at the synchrotron, but staying still is not an option, and as Diamond grows, so does our understanding of the biological chemicals around us.



A design for life

Proteins are the most diverse group of biological molecules, able to perform a dizzying number of functions within living systems. Lynne Regan and colleagues from Yale University explain how biochemists can design new ones to perform specific functions for medicine and research.

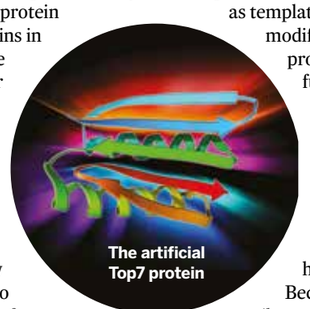
Over billions of years of evolution, proteins have acquired an enormous variety of functions – from tiny chemical messengers such as hormones to the tough molecules that form hair, skin and muscle. Proteins can also act as catalysts for chemical reactions, or form mini-machines that perform specific tasks within cells. Many enable organisms to survive in diverse environments, from scalding deep sea vents to frozen polar caps.

Outside of their cellular context, many of these molecules still carry out their remarkable functions and form the backbone for reagents that advance scientific research and improve patient care in our hospitals. And yet the current range of natural protein functions represents only a fraction of the potential 'protein universe'. Designing proteins in the lab allows us to explore the almost infinite number of combinations of amino acids that it is possible to create.

To explore and expand the protein universe, we use a technique called protein design. The aim is to satisfy two goals simultaneously: to further our understanding of how amino acid sequences affect protein structure and function, and to apply this understanding towards the creation of proteins with high value in research and biotechnological and clinical settings.

There are two main ways to design a protein. The first uses our knowledge of existing protein sequences and structures to design a completely new molecule that has little or no relationship to natural proteins. This is best illustrated by Top7 (see below), a protein composed of typical protein elements, but folded in a way that has never been observed in natural proteins[1].

The second strategy, sometimes referred to as protein redesign, uses natural proteins as templates and makes judicious modifications to alter the protein's structure and/or function[2,3]. A prime example of protein redesign is the 'humanisation' of therapeutic mouse antibodies by grafting the functional components of mouse antibodies onto human antibody scaffolds. Because the bulk of each 'humanised' antibody has a human origin, these antibodies are safer for administration as drugs because of the greatly reduced risk of complications from an immune response.



The artificial Top7 protein

© SCIENCE PHOTO LIBRARY/ALAMY

Given the highly complex relationship between protein sequence and structure, successful execution of both strategies typically requires the testing of multiple sequences to produce a protein that satisfies the design goals. The knowledge gained from careful analysis of successful and failed designs can be applied iteratively to future designs. In other words, we try a bunch of designs, see which ones work and figure out what distinguishes successes from failures.

Of course, the key to protein design is exploring the existing 'protein universe' to understand what makes natural proteins work. Of specific interest is identifying how amino acids in a protein sequence contribute to protein stability and function.

This can help inform a process called consensus design. At the heart of this is the premise that patterns of amino acid conservation in similar proteins – across many species – can be used to identify which amino acids are important in the proteins' structure and function[4,5]. The first step is the alignment of the homologous protein sequences to identify which amino acids have been conserved throughout evolutionary history (Figure 1). For a group of proteins that have similar structures but diverse functions, the conserved amino acids can be interpreted as playing key roles in the protein structure. The amino acids that play functional roles should show poor conservation.

By taking the most frequent amino acid at each position in the alignment, we can create an entire consensus protein that will nearly always be more stable and will have the same structure as the aligned proteins, but will lack function. These consensus proteins make excellent frameworks on which a wide range of functions can be accommodated later. A nice example is the use of tetratricopeptide proteins in cancer treatment[6] (see Outcompeting cancer, page 16).

The boundaries of the protein universe have traditionally been defined by the same 20 amino acid building blocks that make up all natural proteins. However, advances in our understanding of how proteins are made and genome engineering have allowed scientists to break through these boundaries[7].

For designing proteins with non-standard amino acids, we introduce a special codon in the protein-coding mRNA, and reprogram the cell to insert a non-standard amino acid in response to that codon. This approach is particularly useful for protein design because it allows researchers to specify the non-standard amino acid and its position in the protein based on the placement of the hijacked stop codon in the protein coding gene.

Figure 1 (below): Consensus proteins are designed by aligning the sequences of homologous proteins from different organisms. Each row represents one protein, with its amino acid sequence represented by letters A-V. Conservation is indicated by colour, where red indicates absolute conservation across proteins and purple indicates no conservation. The most conserved (ie common) amino acid at each position is used to create a stable protein with the same structure as the natural homologue, to which new functions can be added. In positions where there is no conservation (ie position 2), the amino acid in the consensus protein is arbitrarily chosen.

	1	2	3	4	5	6	7	8
ALLIGATOR	E	K	S	L	K	E	K	G
FRUIT FLY	I	S	S	V	I	A	L	G
BLUE JAY	P	D	L	Q	T	G	L	G
YEAST	S	Q	E	L	A	K	L	G
HUMAN	A	L	E	R	K	N	E	G
RABBIT	A	E	R	L	K	A	E	G
CONSENSUS	A	D	E	L	K	A	L	G

Proteins are particularly suitable for development of advanced drug delivery systems

One example of protein design that uses the above approach is the design of small proteins called TRAPs to interact with specific phosphorylated peptides[3]. Phosphorylation is a very common protein modification and plays a key role in relaying information about the environment to a cell's interior. Because phosphorylation can drastically alter a cell's behaviour, many diseases are associated with improper regulation of phosphorylation signalling. Thus, the ultimate goal of the TRAP designs is to produce a set of proteins, each of which interacts with a specific phosphorylated

Outcompeting cancer

Designing tetratricopeptide repeat (TPR) proteins to fight cancer

Statistical analysis was used to identify regions of TPR proteins that are involved in binding to a major anticancer target known as Hsp90. Hsp90 is a 'chaperone protein' that assists other proteins to fold properly, and is associated with the folding of many cancer associated proteins.

We then grafted the putative binding regions onto a very stable consensus TPR framework. The new protein binds to Hsp90 with high affinity.

This designed TPR protein outcompetes natural Hsp90 co-factors in binding to Hsp90. This competition prevents Hsp90 from carrying out its cellular function.

As a result, levels of HER2, a cancer associated binding partner of Hsp90, fell in HER2-positive breast cancer cells[6]. Similar use of consensus proteins as design templates has been applied to a variety of proteins, simultaneously enhancing our understanding of protein structure and creating novel proteins that can fulfil current therapeutic needs.

Dividing breast cancer cell

peptide, and so provides the means for *in vitro* and *in vivo* detection of aberrant phosphorylation signalling.

How else can we take full advantage of the expanding protein universe? Proteins are particularly suitable for development of advanced drug delivery systems and tissue engineering. Specifically, proteins' diverse functionality and biocompatibility give them an advantage over many synthetic materials in that they can be reabsorbed by the body after use and are less likely to cause complex immune responses.

Our lab has created a new drug delivery system based on 'smart' hydrogels[8,9]. Like jelly, the hydrogels can encapsulate small molecules and respond to mechanical stress like natural tissues. Unlike jelly, however, the bonds that hold our smart hydrogels together are genetically encoded to respond to physiological stimuli, such as a change in pH or ionic strength. These hydrogels dissolve in low pH conditions, such as those observed in cancer microenvironments, and therefore can deliver their encapsulated cargo specifically to cancer cells (Figure 2).

All of the advances above illustrate how protein design is being used to create amazing new molecules that enable us to develop innovative tools for biomedical therapies and also better understand the evolutionary forces that have shaped our existing protein universe. Ongoing efforts have been greatly enhanced by the creation of highly stable consensus proteins and an ever expanding repertoire of amino acid units. Clever assembly of these new proteins into complex hydrogel networks allows the direct translation of relatively simple designed proteins into valuable biomedical materials for tissue engineering, targeted drug delivery and other research and clinical applications.

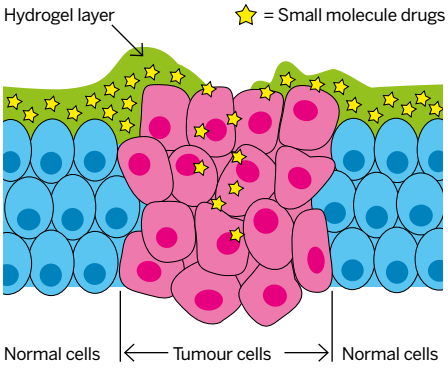


Figure 2: The targeted delivery of small molecule drugs (depicted as yellow stars) to tumour cells (pink) using 'smart' hydrogels. The acidity of the tumour cells causes the hydrogel layer (green) to break down at the tumour interface while staying intact at the normal cell layer (blue), allowing for the controlled release of molecular cargo.

Lynne Regan is professor of molecular biophysics, biochemistry and chemistry at Yale University. The Regan Lab is interested in the relationships between the structure, function and stability of macromolecules.

Danielle Williams is a graduate student in the department of molecular biophysics and biochemistry at Yale University. Her current research involves the design and creation of new protein-based nanomaterials.

Nicholas Sawyer is a graduate student in the department of Molecular Biophysics and Biochemistry at Yale University. His current research focuses on designing novel proteins to detect protein post-translational modifications.

References

1) Kuhlman, B. *et al*. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
2) Jackrel, M. E. *et al*. Redesign of a protein-peptide interaction: characterization and applications. *Protein Sci.* **18**, 762–774 (2009).
3) Sawyer, N. *et al*. Designed phosphoprotein recognition in *Escherichia coli*. *ACS Chem. Biol*. **9**, 2502–2507 (2014).
4) Main, E. R. G. *et al*. Design of stable α-helical Arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).
5) Sawyer, N. *et al*. All repeats are not equal: a module-based approach to guide repeat protein design. *J. Mol. Biol.* **425**, 1826–1838 (2013).
6) Cortajarena, A. L. *et al*. Designed TPR modules as novel anticancer agents. *ACS Chem. Biol*. **3**, 161–166 (2008).
7) Rovner, A. J. *et al*. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89–93 (2015).
8) Grove, T. Z. *et al*. Stimuli-responsive smart gels realized via modular protein design. *J. Am. Chem. Soc*. **132**, 14024–14026 (2010).
9) Grove, T. Z. *et al*. A modular approach to the design of protein-based smart gels. *Biopolymers* **97**, 508–517 (2012).